IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

*INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING*
*Vol. 4, Issue 4, April 2016*

# Speaker Recognition System Using Watermark Technology for Anti-Spoofing Attack: A Review

**Nihalkumar G. Desai[1], Nikunj V. Tahilramani[2]**

Research Scholar, Electronics and Communication Department, C.G.P.I.T., Bardoli, India [1]

Assistant Professor, Electronics and Communication Department, C.G.P.I.T., Bardoli, India [2]

**Abstract**: This paper is about Speaker recognition using Watermark technology for Anti-Spoofing Attack. Speaker recognition is a process whereas speaker identification and speaker verification refer to definite tasks. For the areas in which security is a foremost concern, speaker Recognition technique is one of the most useful recognition techniques. There are possibilities of spoofing attack in recognition system, which break the Security system. By using the watermark technology the authenticity information can be hiding. That hidden information can use as the authenticity or surety purpose in Speaker recognition System.

**Keywords**: Speaker Recognition, Anti-spoofing, watermark technology, speech watermarking.

## I. INTRODUCTION

Speaker recognition system is used for the security purpose, telephone banking and voice mail system, there are many challenges in speaker recognition system, which are affecting directly or indirectly affect the system accuracy. One of that is voice conversion which is known as spoofing attack. In spoofing attack one speaker speech is produce in source side is modified like at target speech [2]. Mostly two popular spoofing attack methods include speech synthesis system and a human mimicking [2]. In speech synthesis between source and target is trained and test or Target speaker speech is directly applied for synthesizer. In human voice mimicking a person try to generate speech like to target speaker or target's speech is recorded and then played back [2]. Although studies have shown that human can easily distinguished between synthesized and natural speech, it is difficult even for human to distinguish play-back attacks.

This paper has used the new idea for anti-spoofing attack in speaker recognition. The main idea is that the watermark is embedded in speech signal at transmitter side can be applied for checking genuinely of the speaker in receiver side. Due to properties of the watermark, various type of spoofing attack can be secured. Furthermore, there is possibility to trace the source of attack. That gives a better authenticity of speaker.

This paper is organized as following: First, speaker recognition in speech, second watermarking in speech is discussed, third application of speech watermarking in anti-spoofing attack is explained. Finally, performance parameters and conclusion are discussed.\

## II. SPEAKER RECOGNITION SYSTEM

Fig. 1 shows the basic model of speaker recognition system. Speaker recognition system can be design using three phases that is pre-processing, feature extraction and classification [3].

Fig. 2 shows block diagram of a speaker recognition system. Microphone is used as a sensor. Sensor data is given to pre-processing block. After finding start point and end point in pre-processing, the features are extracted frame by frame and it will store as template in template database [3]. One by one 1 to N all speaker's speech data are stored in template database. This procedure is called as Enrollment, also called as training phase.

During the testing phase one of the N speakers will speak and this data are given to the pre-processing block then find the features and prepared a template. Now that template will be matched with the template database and the best matched will considered as best score depending upon the threshold. The best among of that decide and identified the true speaker.

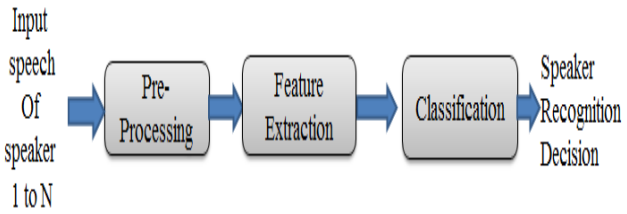


Fig. 1. Basic Model of Speaker Recognition System

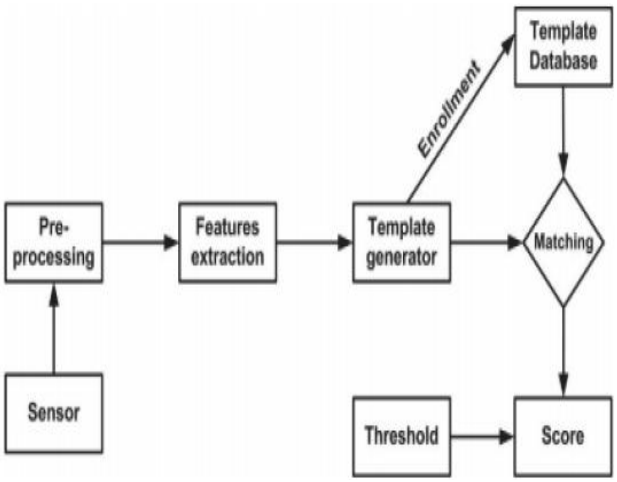


Fig. 2. Block diagram of Speaker recognition process [1]


DOI 10.17148/IJIREEICE.2016.4436

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 4, April 2016

### A. Pre-Processing

The property of speech signal is change relatively slowly with time. So that short time analysis is needed in speech processing. In speech processing this short time segment is considered as frame and the frame size is taken as 10ms to 40ms so that variation of speech signal is observable in short time [3]. Speech is divided in number of frame in which all the frame Short Time Energy (STE) [5] and Zero Crossing Rates (ZCR) [5] is measured. If the energy of any frame is higher than some threshold then it is considered as signal frame. If the energy is less than threshold then it is considered as silent period. So, energy is widely used for the measurement of start and end point of any speech signal. But for weak fricative it is not possible to find the start and end point by simply finding the energy only. ZCR is used for finding the weather the frame is voiced or unvoiced. If the ZCR counts are found to be higher, then it is unvoiced frame and if ZCR counts are less, then it is voiced frame. Also for silent period the ZCR counts are always less than the unvoiced sound. So, Based on this STE and ZCR One can accurately find Start point and end point of any speech signal. Now this speech is applied to the next phase called as feature extraction technique.

### B. Feature Extraction Techniques

- *Mel Frequency Cepstral Coefficient (MFCC)*

A block diagram of an MFCC feature extraction is shown in Fig. 3. This coefficient has a great success in speaker recognition application. The MFCC is the most evident example of a feature set that is extensively used in speech recognition [6]. As the frequency bands are positioned logarithmically in MFCC, it approximates the human system response more closely than any other system. Technique of computing MFCC is based on the short-term analysis, and thus from each frame a MFCC vector is computed. In order to extract the coefficients the speech sample is taken as the input and hamming window is applied to minimize the discontinuities of a signal [7]. Then DFT will be used to generate the Mel filter bank.

MFCC can be computed by using the below formula 1.

$$\text{Mel(f)} = 2595 * \log_{10}(1 + \text{f}/700) \qquad (1)$$

The following fig. 3 shows the steps involved in MFCC feature extraction.

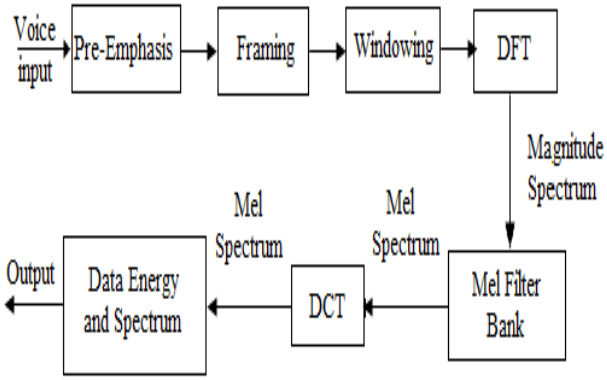


Fig. 3. Block diagram of Mel frequency cepstral coefficient[7]

- *Linear Predictive Coding (LPC)*

Linear prediction is a mathematical computational operation which is linear combination of several previous samples. LPC of speech has become the predominant technique for estimating the basic parameters of speech. It provides both an accurate estimate of the speech parameters and it is also an efficient computational model of speech.

The basic idea behind LPC is that a speech sample can be approximated as a linear combination of past speech samples [8]. Through minimizing the sum of squared differences (over a finite interval) between the actual speech samples and predicted values, a unique set of parameters or predictor coefficients can be determined. These coefficients form the basis for LPC of speech. The following fig. 4 shows the steps involved in LPC feature extraction.

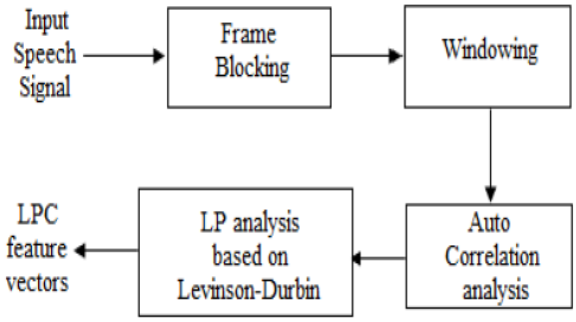


Fig. 4. Block diagram of Linear predictive coding [8]

- *Other feature extraction techniques*

A number of other feature extraction techniques are available by simply modifying the above feature extraction techniques, including the following

- Log area ratio (LAR)[9]
- Jitter and shimmer[9]
- Mean and Variance of the residual phase[9]
- Combination of MFCC and Phase information[12]
- Mel filter bank slope (MFS) features[10]
- Delta and double delta of MFCC features[10]
- Cepstral Mean Subtraction (CMS) MFCC[11]

So many feature extraction techniques are available. One can used any technique according to application and can get better recognition accuracy. Now this features we have to store in a database for N different speakers and this database will be used for classification purpose.

### C. Classification Techniques

- *Dynamic Time Warping*

This is used specifically to deal with variance in speaking rate and variable length of input vectors because this algorithm calculates the similarity between two sequences which may vary in time or speed. To normalize the timing differences between test utterance and the reference template, time warping is done non-linearly in time dimension. After time normalization, a time normalized distance is calculated between the patterns. The speaker with minimum time normalized distance is identified as authentic speaker [5].

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 4, Issue 4, April 2016

• *Other Classification Techniques*

A number of classification techniques are available including following:

- Artificial Neural Network (ANN) [3]
- Polynomial Classifier [11]
- Vector quantization [14]
- Gaussian mixture model [14]
- Support Vector Machines [15]
- Hidden Markov Model [16]
- Pearson correlation [17]

## III. SPEECH WATERMARKING

Watermarking is the technique and art of hiding additional data (such as watermarked bits, logo and text message) in the host signal which includes image, video, audio, speech, text, without any perceptibility of the existence of additional information[18]. The additional information which is embedded in the host signal should be extractable and must resist various intentional and unintentional attacks. Digital speech watermarking process is depicted in Fig. 5.

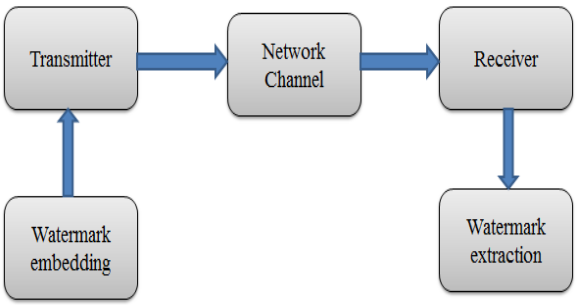


Fig. 5. Fundamental architecture of digital speech watermarking

In terms of source and extraction module for digital speech watermarking are found three main categories [18]:

- Blind [23] speech watermarking which does not need any extra information such as original signal, logo or watermarked bits for watermark extraction.
- Semi-blind speech watermarking which may need extra information for the extraction phase like access to the published watermarked signal that is the original signal after just adding the watermark.
- Non-blind speech watermarking which needs the original signal and the watermarked signal for extracting watermark.

Different methods are used for digital speech watermarking [18]. Fig. 6 presents an overview of these methods. There is Steganography [4][20][21] Techniques also an option for information hiding in Audio and speech signal.

## IV. WATERMARKING FOR ANTI-SPOOFING ATTACK

There are possibility of spoofing and attack in the speaker recognition system such like whenever the input side or sensor side the claim speaker data is already know the watermark of the system so that playback attack is possible in input side which make spoof to the system. There is also possible to attack at the transmission line with replay attack or direct attack to the system [19]. Figure 7 is show the possibility of the attack in the system. To prevent the system from the attack we can use the watermark technology at transmitter side as well as receiver side.

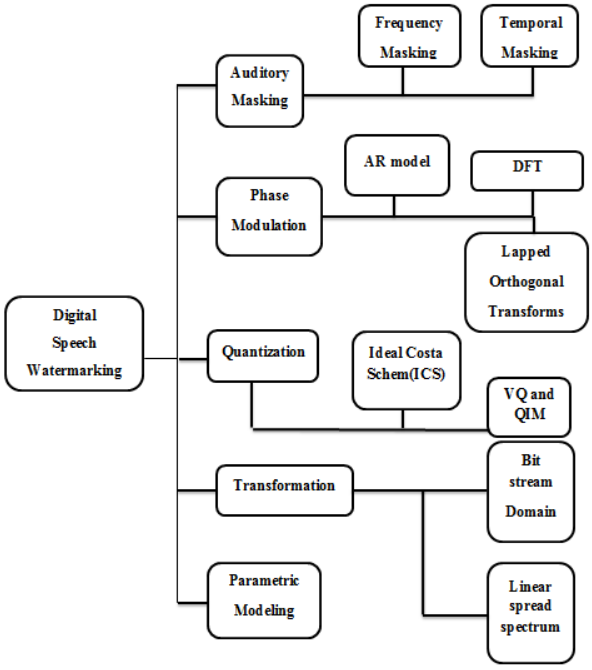


Fig. 6. An overview of different methods for digital speech watermarking[18]

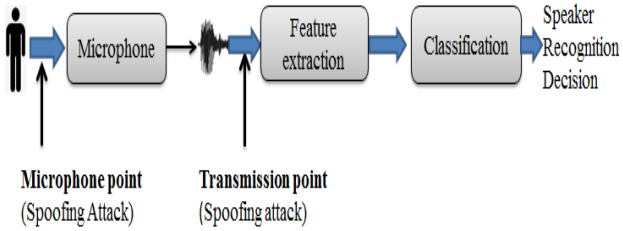


Fig. 7. Possible Spoofing Attack in Speaker recognition System.

Due to nice property of digital speech watermarking for authentication, it is possible to authentication or verify genuinely of the speaker in receiver side. Fig. 8 shows proposed system in transmitter side. As seen, first the speech signal is checked for available watermark, if watermark is available in speech signal, it means the signal is already has been used (replay attack) so that the speaker is unauthorized. Otherwise watermark is embedded in speech signal as anti-spoofing attack.

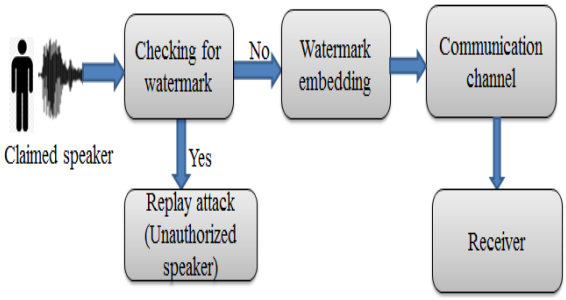


Fig. 8. Anti-spoofing attack by using digital speech watermarking at tranmiter.

Fig. 9 shows that in receiver side, as seen firstly speakers feature is extracted from claimed speech, secondly this feature is feed to classification model then decision is made. Whole of this process is same as normal speaker recognition. However when the speaker is accepted, it would be checked for watermark availability. If watermark is available in claimed speech, it means the source of the claimed speaker is genuine. Otherwise, spoofing attack is happened. Sometimes, it is possible to track the replay attack for finding the source of attack [2].

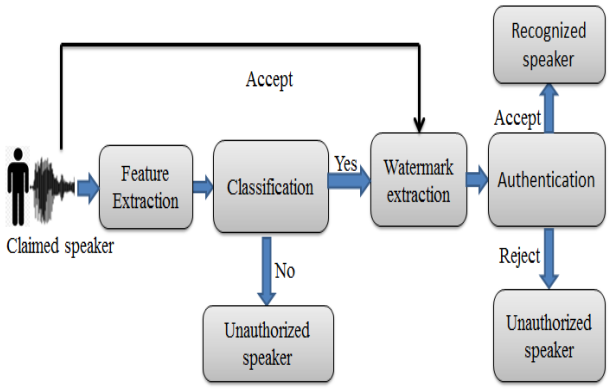


Fig. 9 Diagram of speaker recognition with an anti-spoofing attack detector by using digital speech watermarking

## V. PERFORMANCE MEASURING PARAMETERS

### A. Identification Rate

Identification Rate is familiar measurement of the performance of a speaker recognition system [3].

$$\%\text{Identification Rate} = \frac{No.Of\ Correctly\ \ Identified\ \ trials}{Total\ \ No.of\ Trails} \quad (2)$$

### B. Signal to Watermark Ratio

Signal to watermark ratio is investigating the effect of the watermark on speaker recognition system [2].

$$SWR(\omega,\acute{\omega})=10log_{10}\frac{\sum_{i=1}^{N}\omega(i)^2}{\sum_{i=1}^{N}[\omega(i)-\acute{\omega}(i)]^2}(\text{dB}) \quad (3)$$

Where $\omega$ and $\acute{\omega}$ are original and watermarked speech signal respectively.

According watermarking method is use, related performance parameters are change [18] [22].

## VI. CONCLUSION

Spoofing attacks are the main aim for developers who want to develop remote or online speaker recognition system. Digital watermarking can successfully use for various types of spoofing attack.it improve accuracy of speaker recognition system in case of unsecure channels.

## REFERENCES

[1] Alfredo Maesa, Fabio Garzia, Michele Scarpiniti and Roberto Cusani, "Text Independent Automatic Speaker Recognition System Using Mel- Frequency CepstrumCoefficient and Gaussian Mixture Models," Journal of Information Security, Volume 3, Issue 4, October, 2012, pp.335-340.

[2] M.A Nematollahi, S.A.R Al-Haddad, Shyamala Doraisamy and M. Ranjbari, "Digital Speech Watermarking for Anti-Spoofing Attack in Speaker Recognition", IEEE Region 10 Symposium, 2014, pp. 476-479.

[3] Kinnal Dhameliya and Ninad Bhatt, "Feature Extraction and Classification Techniques for Speaker Recognition: A Review", IEEE international Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), Visakhapatnam, January 2015.

[4] Nikunj.V.Tahilramani and Ninad Bhatt, "Steganography in Speech Signal with Enhanced Multi-Pulse Excitation Codevector with Reduced Number of Bits", IEEE January, 2015.

[5] Nidhi Desai, Kinnal Dhameliya and Vijayendra Desai, "Recognizing voice commands for robot using MFCC and DTW," International Journal of Advanced Research in Computer and Communication Engineering, Volume 3, Issue 5, May, 2014.

[6] Vibha Tiwari, "MFCC and its applications in speaker recognition", International Journal on Emerging Technologies, 2010, pp.19-22.

[7] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques", journal of computing, Volume 2, Issue 3, March 2010 pp. 137-143.

[8] Om Prakash Prabhakar and Navneet Kumar Sahu, "A Survey On: Voice Command Recognition Technique", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 5, May, 2013, pp. 576-585.

[9] Jianglin Wang, An Ji and Michael T. Johnson, "Features for Phoneme Independent Speaker Identification," IEEE International Conference on Audio Language and Image Processing (ICALIP), Shanghai, July, 2012, pp. 1141-1145.

[10] Srikanth R Madikeri and Hema A Murthy, "Mel Filter Bank Energy- Based Slope Feature and Its Application to Speaker Recognition," IEEE National Conference on communication (NCC), Bangalore, January, 2011, pp. 1-4.

[11] Hemant A. Patil, Purushotam G. Radadia and T. K. Basu, "Combining Evidences from Mel Cepstral Features and Cepstral Mean Subtracted Features for Singer Identification," IEEE International Conference on Asian Language Processing, Hanoi, November, 2012, pp. 145-148.

[12] Seiichi Nakagawa, Longbiao Wang and Shinji Ohtsuka, "Speaker Identification and Verification by Combining MFCC and Phase Information," IEEE transaction on audio, speech and language processing, Volume 20, Issue 4, May, 2012, pp. 1085-1095.

[13] Dr E.Chandra, K.Manikandan and M.S.Kalaivani, "A Study on Speaker Recognition System and Pattern classification Techniques", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 2, Issue 2, February, 2014, pp. 963-967.

[14] R. P. Ramachandran, K.R. Farrell, R. Ramachandran and R. J. Mammone, "Speaker Recognition—General Classifier Approaches and Data Fusion Methods", Pattern Recognition in Information Systems, Volume 35, Issue-12, December, 2002, pp. 2801-2821.

[15] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery (Springer), Volume 2, Issue-2, June, 1998, pp. 121-167.

[16] Ghahramani Z., "An Introduction to Hidden Markov Models and Bayesian Networks", International Journal of Pattern Recognition and Artificial Intelligence, Volume. 5, Issue-1, 2001, pp. 9-42.

[17] Mayur R Gamit, and Kinnal Dhameliy. "English Digits Recognition using MFCC, LPC and Pearson's Correlation", International Journal of Emerging Technology and Advanced Engineering, Volume 5, Issue 5, May 2015, pp. 364-367.

[18] Nematollahi, Mohammad Ali, and S. A. R. Al-Haddad. "An overview of digital speech watermarking." International Journal of Speech Technology, 2013, pp. 1-18.

[19] ZhizhengWua,, Nicholas Evansb, Tomi Kinnunen, Junichi Yamagishid, Federico Alegreb, and Haizhou Lia, "Spoofing and countermeasures for speaker verification: a survey", Speech communication, September, 2014.

[20] Nikita A. Malhotra and Nikunj Tahilramani, "Steganography Approach of Weighted Speech Analysis with and without Vector Quantization using Variation in Weight Factor", International Journal of Current Engineering and Technology, Volume 3, Issue 3, June, 2014, pp. 1334-1336.

[21] Fatiha Djebbar, Beghdad Ayad Karim , Abed Meraim and Habib Hamam, Comparative Study of Digital Audio Steganography Techniques," EURASIP journal on audio, speech and music processing, springer, 2012.

[22] Seethal Paul and Sreelakshmi T.G, "Performance Analysis and Study of Audio WatermarkingAlgorithms", International Journal Of

Engineering And Computer Science, Volume 3, Issue 8 August, 2014, pp. 7540-7547.

[23] Nematollahi, Mohammad Ali, S. A. R. Al-Haddad, Shyamala Doraisamy, F. Zarafshan. "Blind Digital Speech watermarking Based on Eigen-value Quantization In DWT." Journal of King Saud University, December, 2014, pp. 58-67.

[24] Faundez-Zanuy, Marcos, Martin Hagmüller, and Gernot Kubin. "Speaker identification security improvement by means of speech watermarking." Pattern recognition, February, 2007, pp.3027-3034.

[25] J.P. Campbell Jr, "Speaker recognition: A tutorial," Proceedings of the IEEE, Volume. 85, Issue. 9, 1997, pp. 1437–1462.

[26] Wu, Zhizheng, et al. "A study on spoofing attack in state-of-the-art speaker verification: the telephone speech case" Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC) Asia-Pacific IEEE, 2012.

## BIOGRAPHIES

**Nihalkumar G. Desai** received his B.E., from S.S.A.S.I.T, Surat in 2014 and pursuing M.Tech.ICT in Chhotubhai Gopalbhai Patel Institute of Technology (CGPIT), Bardoli. He is working as a Research Scholar, Electronics and communication department, C.G. Patel Institute of Technology. His research interests are antenna designing and speech signal processing.

**Mr. Nikunj Tahilramani** has received B.E from Veer Narmad South Gujarat University, Surat in 2005 and received M.E from Mumbai University in 2012. He is pursuing his PhD degree in the area of Information Hiding in the Speech Signal from Uka Tarsadia University. He has totally 4 years of teaching experience. Presently he is working as an Assistant Professor, Department of Electronics and Communication in Chhotubhai Gopalbhai Patel Institute of Technology (CGPIT), Bardoli. His research interest lies in Speech Processing Applications, Embedded Systems, Signal and Systems, Microprocessors & Microcontrollers and Industrial Automation. He is an active and Life member of IEEE, ISTE and IEI.